Installation of Standalone BLAST on Windows PC

A suite of programs for performing BLAST sequence alignment locally against custom datasets

https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Overview

The Basic Local Alignment Search Tool (**BLAST®**), first introduced to the public in 1990, refers to a collection of sequence alignment programs from NCBI that uses the same heuristic approach to identify the best local alignments between the input query and sequences in the target database. This suite of programs underwent a major change (BLAST 2.0) in 1997 to enable gapped alignment. Changes in 2008 involved the switch the code base from C to C++, leading to the creation of the blast+ package.

NCBI makes the BLAST service, this suite of programs and many target databases, freely available from the BLAST homepage (https://blast.ncbi.nlm.nih.gov/). The latest release of the BLAST package along with a common set of databases are also available through the BLAST FTP site (https://ftp.ncbi.nlm.nih.gov/blast/). You can use this package to perform sequence alignments on you own machine, and more importantly against your own custom datasets. This handout provides an overview of the BLAST package and an outline of the installation and configuration process to ensure smooth execution of needed searches.

Reasons for installing the standalone BLAST+ package

A few example scenarios for setting up the BLAST+ package locally are:

- Large batches of searches and searches that are computationally intensive, such as searches involving dynamic translation of nucleotide sequences (blastx, tblastn, or tblastx)
- Target datasets that are not available at NCBI
- BLAST only serves as an intermediate step within a complex workflow that needs to be automated
- Searches requiring customization that cannot be satisfied by the BLAST web service

Getting the standalone BLAST+ package and BLAST databases

The latest release of standalone BLAST+ packages for common platforms is in the **executables** directory:

> https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/

A set of common BLAST databases is available under the **db** directory:

> https://ftp.ncbi.nlm.nih.gov/blast/db/

For faster download, please access the above through NCBI's Aspera site (Aspera plugin required):

> https://ww.ncbi.nlm.nih.gov/public/?blast/

NCBI provides pre-compiled BLAST+ packages for 64-bit Linux, Mac, and PC platforms. The table below lists the available BLAST+ pages available and their intended platforms. Subsequent sections focus on the installation of binary packages for Windows (in red).

BLAST+ Packages	Chipset	Intended operating system
ncbi-blast-#.#.#+.x86_64.rpm	Intel 64-bit	64-bit linux rpm package
ncbi-blast-#.#.#+-x64-macosx.tar.gz	Intel	MacOSX, equivalent to .dmg
ncbi-blast-#.#.#+-win64.exe	**Intel 64-bit**	**64-bit Windows, equivalent to win64.tar.gz**
ncbi-blast-#.#.#+-x64-linux.tar.gz	Intel 64-bit	64-bit Linux
ncbi-blast-#.#.#+-x64-win64.tar.gz	Intel 64-bit	64-bit Windows, equivalent to win64.exe
ncbi-blast-#.#.#+.dmg	Intel	MacOSX disk image, equivalent to maxosx.tar.gz

NOTE:
1. The above table omits the BLAST+ source code files in .gz, zip, and dmg formats.
2. NCBI no longer offers pre-compiled packages for 32-bit platforms.
3. The -remote option requires HTTPS protocol, this option only works with release 2.4.0 or later.
4. Resources are limited and NCBI can only support the most current release.

## Steps for downloading and installing a selected BLAST+ package

The installation procedures for Windows PC and Linux/Unix-like systems are very different. For installation on Linux/Unix platforms see relevant document at https://www.ncbi.nlm.nih.gov/books/NBK1762/. This handout describes the steps needed to install the package on PCs running the Microsoft Windows operating system.

Index of /blast/executables/blast+/LATEST

Name	Size	Date Modified
[parent directory]		
ChangeLog	85 B	10/18/17, 8:00:00 PM
ncbi-blast-2.7.1+-1.src.rpm	18.1 MB	10/18/17, 8:00:00 PM
ncbi-blast-2.7.1+-1.src.rpm.md5	62 B	10/18/17, 8:00:00 PM
ncbi-blast-2.7.1+-1.x86_64.rpm	181 MB	10/18/17, 8:00:00 PM
ncbi-blast-2.7.1+-1.x86_64.rpm.md5	65 B	
ncbi-blast-2.7.1+-src.tar.gz	22.2 MB	
ncbi-blast-2.7.1+-src.tar.gz.n		
ncbi-blast-2.7.1+-src.zip		
ncbi-blast-2.7.1+-src.zip.md		
ncbi-blast-2.7.1+-win64.exe		
ncbi-blast-2.7.1+-win64.exe.		
ncbi-blast-2.7.1+-x64-linux.t		
ncbi-blast-2.7.1+-x64-linux.tar.gz.md5	69 B	
ncbi-blast-2.7.1+-x64-macosx.tar.gz	127 MB	

The example below uses version 2.7.1, ncbi-blast-2.7.1+-win64.exe. Installation of a newer version changes the directory name, so you need to take the name change into consideration in the subsequent configuration steps.

Go to https://ftp.ncbi.nlm.nih.gov/blast/executables/LASTEST/ using a browser (**A**). Place the mouse pointer over the file with "win64.exe" in its name, right click on the file, and select "Save Link As …" (**B**). For easy file access, save the file to the Desktop (**C**).
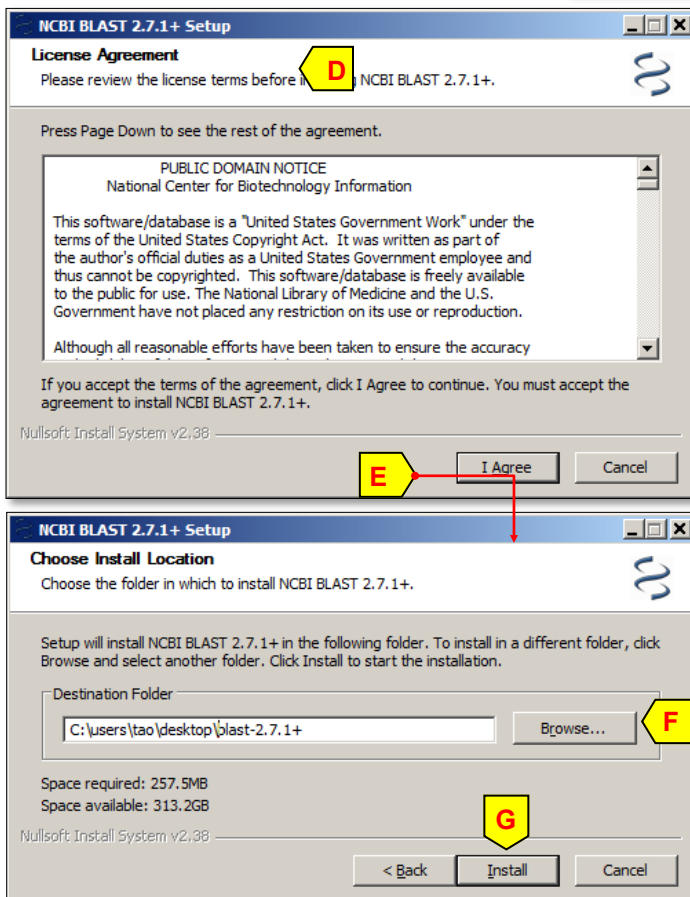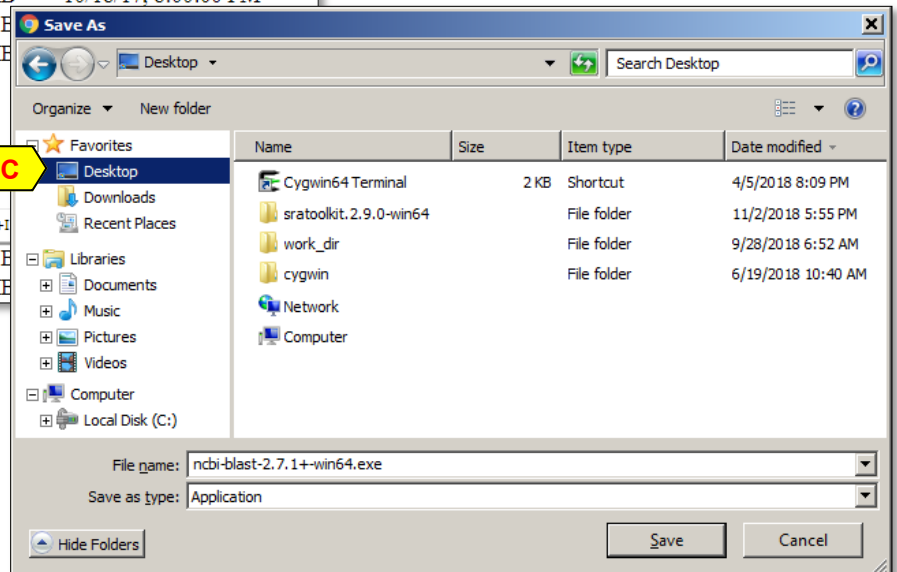
Double click the downloaded file on the Desktop to launch the installer (**D**). Click the "I Agree" button to see a new prompt (**E**) and customize the target directory by editing the text or using the browse button (**F**). The example set it to "C:\users\tao\desktop\blast-2.7.1+". Your own directory structure (red) will be different. If you select another destination, make sure you avoid destinations with space-containing names. Complete the process by clicking the "Install" button (**G**). This also appends the path to the bin directory under blast-2.7.1+ to the existing value of the **Path** environment variable.

Making BLAST databases from FASTA files
The recommended steps are:
- Create a **db** directory under the blast-2.7.1+, as part of the configuration procedures
- Configure BLASTDB environment variable to point to this newly created db directory
- place the FASTA file in the db directory
- cd to the db directory
- run the **makeblastdb** program to convert the FASTA file into a BLAST database

We will describe configuration (p.3) and provide example command lines for the last two steps (p.4).

Configuring the installation

Create a subdirectory to store database files at a central location

Open the blast-2.7.1+ directory in Window Explorer, create a new directory there (right-clicking and selecting "New >> Folder"), then rename it to "db" (**A**).
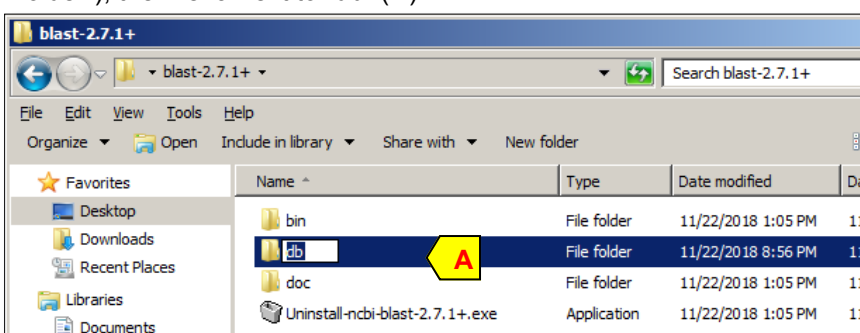
Create an environment variable to inform other BLAST+ programs

BLAST+ programs read from the BLASTDB environment variable to locate the user-specified databases. To create this environment variable, launch the "Control Panel" from the Start menu. Click the System icon, then the "Advanced system settings" link to open the "System Properties" prompt (**B**). Click the "Environment Variable …" button (**C**) under the "Advanced" tab to access the "Environment Variables" dialog box (**D**). The highlighted entry is the user variable path, which contains a "C:\users\ttao\desktop\blast+2.7.1+\bin" for the operating system to locate the blast+ programs (details not shown). Click the "New" button to create a new one named BLASTDB and specify its value as indicated (**E**).
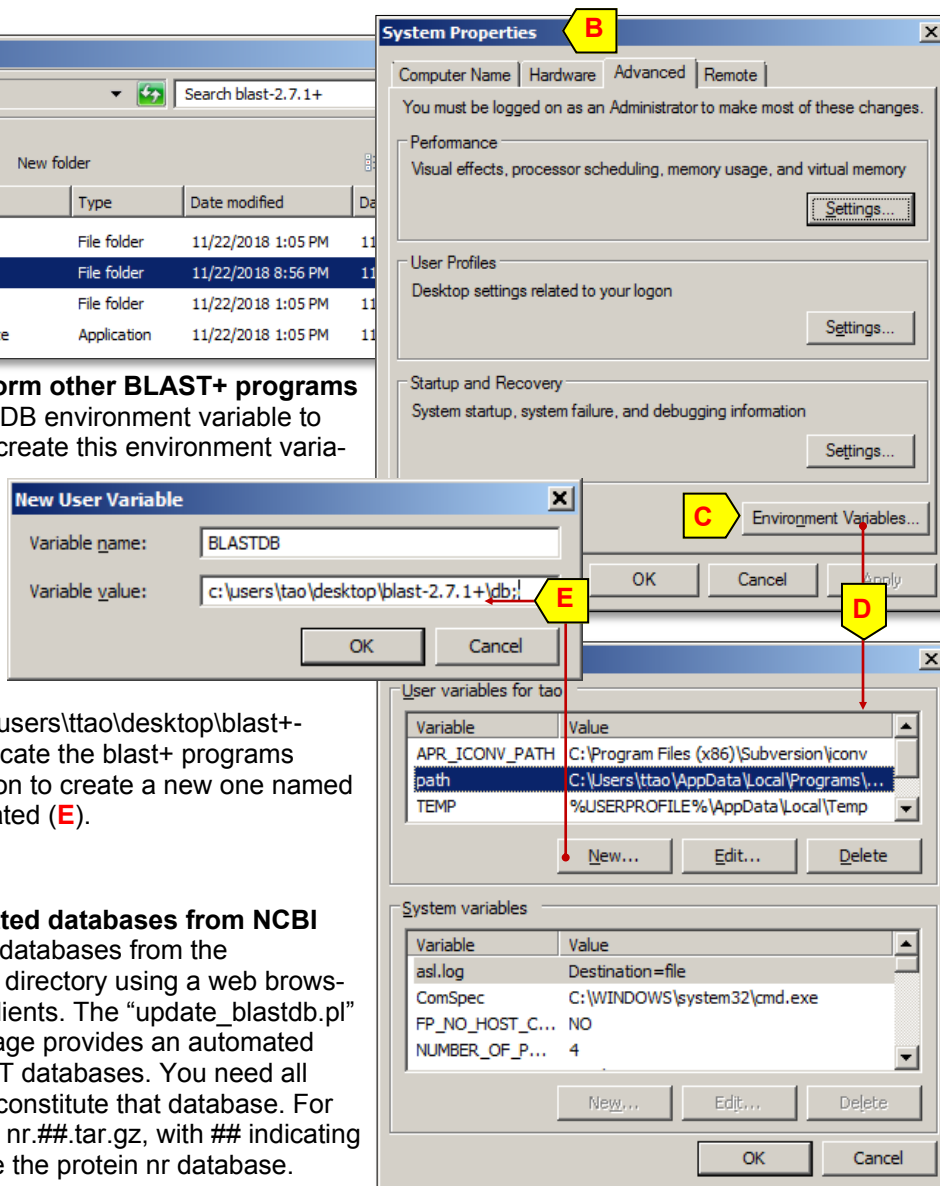
Populating the db folder

Downloading and installing preformatted databases from NCBI

You can download preformatted BLAST databases from the https://ftp.ncbi.nlm.nih.gov/blast/db/ FTP directory using a web browser, the Aspera browser plug-in, or FTP clients. The "update_blastdb.pl" Perl script included in the BLAST+ package provides an automated method to download and maintain BLAST databases. You need all volumes with the same base-name to reconstitute that database. For example, you need to get all files named nr.##.tar.gz, with ## indicating different volume numbers, to reconstitute the protein nr database.
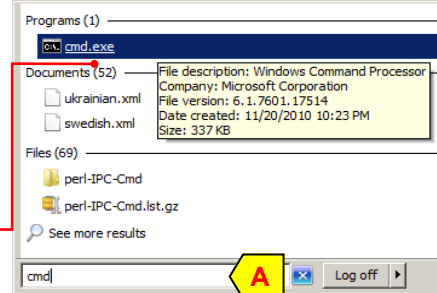

You can download the database manually through web browsers. For example, to download the second volume of the refseq_rna database (refseq_rna.01.tar.gz), right click the file name and select "Save Link As …" (**F**), then select to save the file to the newly created **db** folder (**G**) in the prompt (with the first volume already downloaded).
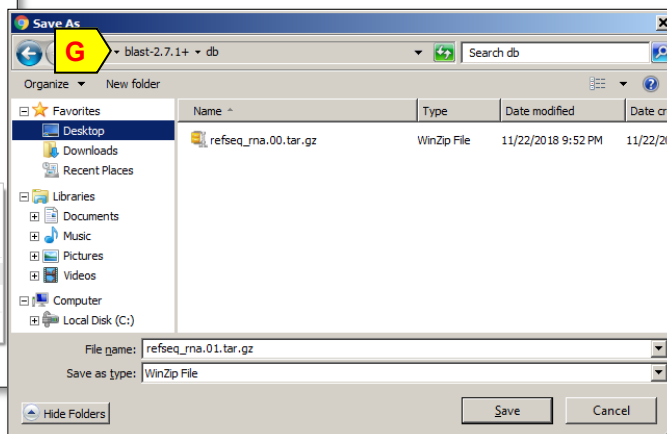
Expanding and extracting the downloaded files with WinZip or other decompressing utilities creates the ready to use database. Multi-volume databases will contain an alias file that ties the individual volumes into a single virtual database. This alias file also makes it possible to address all volumes of the database in a single call.

Testing the installation

A BLAST search contains several components: the installed program, a database, and a file with query sequence(s). In standalone setup, you need issue instructions through a command prompt since BLAST programs do not have graphical user interfaces. You can locate the command prompt by searching for cmd (**A**).

```
C:\ Windows Command Processor                                                    _ □ ×
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Windows\System32> cd c:\users\tao\desktop\blast-2.7.1+

c:\Users\tao\Desktop\blast-2.7.1+> dir
 Volume in drive C is Local Disk
 Volume Serial Number is 9ECC-F14A

 Directory of c:\Users\tao\Desktop\blast-2.7.1+

11/22/2018  10:32 PM    <DIR>          .
11/22/2018  10:32 PM    <DIR>          ..
11/22/2018  01:05 PM    <DIR>          bin
11/22/2018  10:27 PM    <DIR>          db
11/22/2018  01:05 PM    <DIR>          doc
11/22/2018  10:40 PM             2,807 test_query.fna
11/22/2018  01:05 PM            62,561 Uninstall-ncbi-blast-2.7.1+.exe
               2 File(s)         65,368 bytes
               5 Dir(s)  329,711,644,672 bytes free

c:\Users\tao\Desktop\blast-2.7.1+> blastdbcmd -db refseq_rna.00 -info
Database: NCBI Transcript Reference Sequences
        2,426,648 sequences; 3,992,088,684 total bases

Date: Nov 21, 2018  3:00 AM     Longest sequence: 108,067 bases

Volumes:
        c:\users\tao\desktop\blast-2.7.1+\db\refseq_rna.00

c:\Users\tao\Desktop\blast-2.7.1+> blastdbcmd -db refseq_rna.00 -entry NM_000249
>test_query.fna

c:\Users\tao\Desktop\blast-2.7.1+> blastn -db refseq_rna.01 -query test_query.fna
-outfmt 6 -max_target_seqs 5
NM_000249.3  XM_001170433.3  99.587  2666    7       1       1       2662
1       2666    0.0     4859
NM_000249.3  XM_003826110.2  99.587  2666            1       1       2662
14      2679    0.0     4859
NM_000249.3  XM_007971803.1  97.661  2651    56      4       1       2646
25      2674    0.0     4547
NM_000249.3  XM_007971804.1  97.586  2651    55      5       1       2646
25      2671    0.0     4532
NM_000249.3  XM_010376619.1  96.523  2560    67      4       111     2662
1       2546    0.0     4215

c:\Users\tao\Desktop\blast-2.7.1+> blastn -help |more
USAGE
  blastn [-h] [-help] [-import_search_strategy filename]
    [-export_search_strategy filename] [-task task_name] [-db database_name]

DESCRIPTION
  Nucleotide-Nucleotide BLAST 2.7.1+

OPTIONAL ARGUMENTS
  -h
   Print USAGE and DESCRIPTION;  ignore all other parameters
  -help
   Print USAGE, DESCRIPTION and ARGUMENTS; ignore all other parameters
  -version
   Print version number;  ignore other arguments

c:\Users\tao\Desktop\blast-2.7.1+> cd db
c:\Users\tao\Desktop\blast-2.7.1+\db> makeblastdb -in ..\test_query.fna -dbtype
nucl -parse_seqids -title "demo db with single refseq_rna" -out example_nucl_db

Building a new DB, current time: 11/23/2018 10:36:42
New DB name:   c:\Users\tao\Desktop\blast-2.7.1+\db\example_nucl_db
New DB title:  demo db with single refseq_rna
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.0101983 seconds.
c:\Users\tao\Desktop\blast-2.7.1+\db> dir /D example*
 Volume in drive C is Local Disk
 Volume Serial Number is 9ECC-F14A

 Directory of c:\Users\tao\Desktop\blast-2.7.1+\db

example_nucl_db.nhr   example_nucl_db.nsd
example_nucl_db.nin   example_nucl_db.nsi
example_nucl_db.nog   example_nucl_db.nsq
               6 File(s)          1,026 bytes
               0 Dir(s)  329,659,924,480 bytes free
c:\Users\tao\Desktop\blast-2.7.1+\db> blastdbcmd -db example_nucl_db -info
Database: demo db with single refseq_rna
        1 sequences; 2,662 total bases
Date: Nov 23, 2018  10:36 AM    Longest sequence: 2,662 bases
Volumes:
        c:\Users\tao\Desktop\blast-2.7.1+\db\example_nucl_db
```

Green: input to -title is reflected in the -info output
Yellow: input to -out is reflected in the base database name

In the Command prompt window (**B**), you type instructions (boxed in red) at the cursor:

'cd C:\users\ttao\desktop\blast-2.7.1+' changes the working directory to where the blast package is installed.

'dir' displays the content of this directory.

'blastdbcmd -db refseq_rna.00 -info' checks a volume of installed database.

'blastdbcmd -db refseq_rna.00 -entry NM_000249 > test_query.fna' retrieves NM_000249 from the database in FASTA format (known to exist in this volume) for use as a test query.

'blastn -query test_query.fna -db refseq_rna.01 -outfmt 6 -max_target_seq 5' calls blastn to search test_query.fna against refseq_rna.01 database, requesting tabular format and 5 matches.

'blastn -help | more' displays online help, one screenful at a time.

As mentioned in page 3, the second block of commands are necessary for creating custom databases from input FASTA files. Their functions are explained below.

'cd db' changes working directory from blast-2.7.1+ to the db subdirectory.

'makeblastdb -in ..\test_query.fna -dbtype nucl -parse_seqids -title "demo db with single refseq_rna" -out example_nucl_db' makes a blast database from the test_query.fna FASTA input file in the parental directory, indicates its sequence type as nucleotide, provides a descriptive title, and gives the output with a clear base name pf example_nucl_db. The '-parse_seqids' indexes the firs string in the defline to allow specific sequence retrieval.

'dir /D example*' examines the generated database files.

'blastdbcmd -db example_nucl_db -info' get the summary information of this newly generated database.

Documents and technical assistance

Links to different help documents are online at:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs
More information on standalone BLAST is available from the BLAST user manual:
https://www.ncbi.nlm.nih.gov/books/NBK1762/
Please send your comments, questions, bug reports, and assistance requests to:
blast-help@ncbi.nlm.nih.gov